Investigating dialectal differences using articulography

Martijn Wieling[a,], Fabian Tomaschek[b], Denis Arnold[b], Mark Tiede[c], Franziska Bröker[b], Samuel Thiele[b], Simon N. Wood[d], and R. Harald Baayen[b,e]*

[a]Department of Humanities Computing, University of Groningen, [b]Department of Quantitative Linguistics, University of Tübingen, [c]Haskins Laboratories, [d]Department of Statistics, University of Bath, [e]Department of Linguistics, University of Alberta

**Corresponding author: Martijn Wieling, Oude Kijk in 't Jatstraat, 9712 EK Groningen, Netherlands, +31503635979, wieling@gmail.com*

Abstract
The present study introduces electromagnetic articulography, the measurement of the position of tongue and lips during speech, as a promising method for the study of dialect variation. By using generalized additive modeling to analyze the articulatory trajectories, we are able to reliably detect aggregate group differences, while simultaneously taking into account the individual variation of dozens of speakers. Our results show that two Dutch dialects show clear differences in their articulatory settings, with generally a more anterior tongue position in the dialect from Ubbergen in the southern half of the Netherlands than in the dialect of Ter Apel in the northern half of the Netherlands. This clearly demonstrates that articulography is able to reveal interesting structural differences between dialects which are not visible when only focusing on the acoustic signal.

Keywords: Articulography; Dialectology; Generalized additive modeling; Articulatory setting

Introduction

At present, most studies in dialectology and sociolinguistics investigating pronunciation variation focus on the acoustic properties of vowels (e.g., Clopper & Pisoni, 2004; Labov, 1980; Leinonen, 2010; Recasens & Espinosa, 2005; Adank et al., 2007; Van der Harst, Van Velde & Van Hout, 2014). Since the seminal study of Peterson & Barney (1952), formant measurements have been the method of choice for measuring vowel quality. While the first and second formant are generally assumed to model height and frontness of the tongue body, this relationship is not perfect (Rosner and Pickering, 1994).

Labov, Yaeger and Steiner (1972) have spearheaded the formant-based approach in sociolinguistics by studying English formant-based vowel variation for a large number of speakers from various areas in the United States of America. Since then many other studies assessing dialect variation have used formant-based methods, for example Adank et al. (2007) investigating regional Dutch dialect variation, and Clopper and Paolillo (2006) and Labov, Ash and Boberg (2005) who studied American English regional variation. While formant-based measures provide a convenient quantification of the acoustic signal, the approach is not without its problems. First, since the shape of the vocal tract influences the formant frequencies (e.g., women generally have higher formant frequencies than men), some kind of normalization is required (see Adank et al., 2004 for an overview of various approaches) and choosing one method over another introduces a degree of subjectivity into the analysis. Furthermore, automatic formant detection is far from perfect and requires manual correction in about 17-25% of the cases (Adank et al., 2004; Eklund & Traunmüller, 1997; Van der Harst et al., 2014). Especially when using multiple formant measurement points per vowel (which is arguably better than using only the mid-point of the vowel; see Van der Harst et al., 2014), this becomes very time-consuming. For this reason whole-spectrum methods (obtained by band-pass filtering the complete acoustic signal) have also been used in language variation research. In her dissertation, Leinonen (2010) studied Swedish dialect variation based on the automatic whole-spectrum analysis of Swedish vowel pronunciations. A drawback of this type of analysis, however, is that it is highly sensitive to the amount of noise in the acoustic recordings (Leinonen, 2010, p. 152). Furthermore, both formant-based and whole-spectrum-based methods are not suitable for investigating variation in the pronunciation of consonants.

Another approach to investigating pronunciation variation is to use the transcriptions of the underlying speech signal. By using transcriptions, an abstraction of the acoustic signal is obtained which can be used to assess pronunciation differences between groups of speakers. Even though

"[t]ranscription is a messy thing" (Kerswill & Wright, 1990, p. 273), transcriptions are frequently used in dialectometry where aggregate analyses based on a large set of linguistic items are instrumental for obtaining an objective view of dialectal variation and its social, geographical and lexical determinants (see Wieling and Nerbonne, 2015 for an overview). Obviously, a clear drawback of using transcriptions is that the speech signal is segmented into discrete units, which means that co-articulation effects are generally ignored.

Instead of focusing on transcriptions based on the acoustic signal, it is also possible to examine the articulatory gestures underlying speech (i.e. the movement of lips and tongue, etc. involved in its production; Browman and Goldstein, 1992). Given that ease of articulation is one of the known factors driving linguistic change (Sweet, 1888), this also makes sense from a diachronic perspective. Only a limited number of studies have investigated dialect and sociolinguistic variation by focusing on the movement of the speech articulators. Most of these studies have employed either electropalatography (EPG) or ultrasound tongue imaging. With EPG, the contact between the tongue and the hard palate is monitored with a custom-made speaker-specific artificial palate containing several electrodes. Corneau (2000) applied this method to compare the palatalization gestures in the production of /t/ and /d/ between Belgium French and Québec French, and Recasens and Espinosa (2007) used it to investigate differences in the pronunciation of fricatives and affricates in two variants of Catalan. While EPG only contains information about the tongue's position when it is touching the palate, ultrasound tongue imaging is able to track most of the tongue surface as it moves during the whole utterance. The sociolinguistic relevance of tracking the shape of the tongue was clearly shown by Lawson, Scobbie and Stuart-Smith (2011), who demonstrated that /r/ pronunciation in Scottish English was socially stratified, with middle-class speakers generally using bunched articulations, while working-class speakers more frequently used tongue-tip raised variants. Consequently, Lawson et al. (2011, p.257) suggest that "articulatory data are an essential component in an integrated account of socially-stratified variation."

Unfortunately, there are several drawbacks associated with the two articulatory observational methods described above. The clear drawback of EPG is that it is very costly, as a custom-made artificial palate needs to be constructed for each participant. In addition, EPG does not yield information about the tongue position when it is not touching the palate. While ultrasound tongue imaging does provide this information, it is not always complete as interposed sublingual air pockets are introduced when the tongue is raised or extended, and shadowing from the mandible and hyoid bones may cause the tongue tip and the tongue root to become invisible (Tabain, 2013). Furthermore, analysis of resulting tongue shapes can be impressionistic, as tracking a single flesh point on the tongue is not possible (Lawson et al., 2011; but see Davidson, 2006). Moreover, unless otherwise corrected (cf. Whalen et al. 2005), the imaged tongue shape is relative to the position of the probe and jaw, not to palatal hard structure, and thus evaluation of tongue height across vowels is problematic.

Electromagnetic articulography (EMA; Hoole and Nguyen, 1999; Perkell et al., 1992; Schönle et al., 1987) is a tracking approach which avoids many of these problems. An EMA device tracks as a function of time small sensors attached with dental adhesive to various flesh points within a speaker's vocal tract (e.g., tongue, lips, maxilla, jaw). Radio-frequency transmitters induce voltages in the sensor coils positioned within the field of the device, and sensor position and orientation are subsequently reconstructed by comparing these voltages to known reference values. As a point-tracking technique, it is excellently suited for quantitative analysis. Of course, a clear drawback of the EMA approach is that the tongue cannot be tracked completely. For example, tongue sensors cannot be placed too far back, to prevent triggering the gag reflex of the speaker. Until recently, EMA studies have been conducted with a relatively small number of speakers (e.g., Recasens and Espinosa, 2009: 3 speakers). Because there is much speaker-related variation in articulatory trajectories (Yunusova et al., 2012), it is fortunate that including a larger number of participants is becoming increasingly common (e.g., Yunusova et al., 2012: 19 speakers; Koos et al., 2013: 25 speakers). In our study, we continue this development by including a total of 40 speakers. To our knowledge, this is the largest sample size used in an articulography study to date.

In this study, we focus on Dutch pronunciation variation from an aggregate articulatory perspective. Only a single published study has investigated variation in the Dutch language from an

articulatory perspective.[1] In their study, Scobbie and Sebregts (2010) focused on a single feature, namely allophonic Dutch variation in the pronunciation of /r/ using ultrasound recordings. Unfortunately, due to the low number of speakers (5) and the ultrasound approach, the description of the results remained rather impressionistic.

Of course, many studies have investigated pronunciation variation in Dutch dialects from various other perspectives. For example, as mentioned above, Adank et al. (2007) investigated the acoustic properties of vowels in several regional varieties of Dutch spoken in the Netherlands and Flanders. They observed clear regional variation in the formant-based measurements. Another type of study focusing on Dutch dialects is exemplified by Goeman (1999), who investigated a specific feature in Dutch dialects, namely the loss of [t] in final word pronunciation (i.e. t-deletion). He identified several (geographical constrained) groups within the Netherlands exhibiting specific t-deletion patterns. Following Nerbonne et al. (1996), Heeringa (2004) took an aggregate dialectometric perspective and quantified pronunciation differences by focusing on the transcriptions and comparing those using the edit distance measure. On the basis of comparing hundreds of words between hundreds of locations in the Dutch-speaking language area, he was able to identify the major dialect areas of the Netherlands. In his dissertation (Figure 9.7, p. 234), he identified the four main dialect areas as the Frisian dialect area (in the northwest of the Netherlands), the Limburg dialect area (in the southeast of the Netherlands), the Low-Saxon dialect area (in the northeast of the Netherlands) and the Central Dutch dialect area. Similarly, Wieling et al. (2007, 2011) identified relatively comparable dialect areas using a different dataset of Dutch dialects.

As articulatory data is not readily available for Dutch dialects, we collected dialect (and standard Dutch) pronunciations at two different sites. To ensure the dialects were not too similar, we collected our data at one site in the Low-Saxon dialect area (i.e. the village of Ter Apel), and at another site in the Central Dutch dialect area (i.e. the village of Ubbergen). Given that the goal of this study is to assess articulatory (dialect) pronunciation differences from an aggregate perspective, we include many participants and items. In addition, we propose a flexible statistical approach, generalized additive modeling (GAM; Hastie and Tibshirani, 1990; Wood, 2006) for analyzing articulography data. The advantage of using this approach (explained in more detail below) is that it is able to model the nonlinear trajectories of the tongue sensors in multiple dimensions over time, while also taking into account individual variation. As generalized additive modeling is a regression approach, it is excellently suited to assess the influence of the predictors of interest (in our case the contrast between the two groups) on the articulatory trajectories.

While we expect articulatory differences between the two groups of speakers, due to their different dialect background, we do not have a clear hypothesis about the specific characteristics of these differences. In that sense, our study is exploratory. In the following, we will discuss the methods and results obtained in this study.

Articulatory data collection

Our study was conducted on-site in 2013 at two high schools in the Netherlands. The first school "RSG Ter Apel" was located in Ter Apel (in the northern half of the Netherlands, i.e. in the Low Saxon dialect area), while the second school "HAVO Notre Dame des Anges" was located in Ubbergen (in the southern half of the Netherlands, at a distance of about 150 kilometers from Ter Apel, i.e. in the Central Dutch dialect area). At each school data were collected onsite during a single week by two researchers of the University of Tübingen (MW and DA in Ter Apel and MW and FT in Ubbergen). In Ter Apel, 23 speakers participated, but the data of 2 speakers was excluded as it contained tracking inconsistencies due to a malfunction of the reference sensor. Of the remaining 21 speakers (12 male, 9 female), 15 were high school students born between 1994 and 2000. The other 6 participants were adults born between 1939 and 1967. In Ubbergen, 25 high school students

[1]There is one conference proceedings paper investigating Dutch pronunciation variation from an aggregate articulatory perspective (Wieling et al., 2015). However, the present study is an extended version of that study, and offers a more detailed report of the methods and results presented by Wieling et al. (2015). In addition, this study does not only focus on dialect variation, but also on variation in standard Dutch. Note that the results presented here are slightly different from those discussed by Wieling et al. (2015), as in the present study the data was reanalyzed using an improved version of the generalized additive modeling software.

participated, but the data of 6 speakers was excluded (5 speakers did not speak the regional dialect, and the reference sensor malfunctioned for 1 speaker). The remaining 19 participants (17 male, 2 female) were born between 1994 and 2000. Before participating, participants were informed about the nature of the experiment and required to sign the informed consent form. Each data collection session lasted a total of 50 minutes for which the participants were compensated with €10.

The articulography data were collected with a portable 16-channel EMA device (WAVE, Northern Digital) at a sampling rate of 100 Hz, and automatically synchronized to the audio signal (recorded at 22.05 kHz using an Oktava MK012 microphone) by the controlling software (WaveFront, Northern Digital). This software also corrected for head movement using a 6DOF reference sensor attached to each participant's forehead. The microphone and EMA device were connected to the controlling laptop via a Roland Quad-Capture USB Audio interface.

We attached a total of three sensors to the midline of each participant's tongue using PeriAcryl 90 HV dental glue. One sensor was positioned as far backward as possible without causing discomfort for the speaker. Another sensor was positioned about 0.5 cm behind the tongue tip. The final sensor was positioned approximately midway between the other two sensors.[2] Attaching all sensors took about 20 minutes. Whenever sensors came off during the course of the experiment, they were reattached at their original location. To align the positional data to axes comparable between speakers, a separate biteplate recording (containing 3 sensors, see Figure 1) was used during processing to rotate the data of each speaker relative to the occlusal plane (Hoole & Zierdt, 2010; Yunusova et al., 2009) and to translate to a common origin on the biteplate ('X' in Figure 1; note that this origin does not influence the normalized sensor positions, due to our preprocessing steps outlined below).

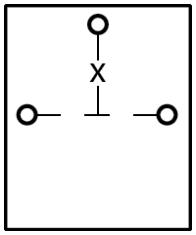

Figure 1. Schematic representation of the biteplate. Circles mark the sensor positions. The 'X' marks the origin.

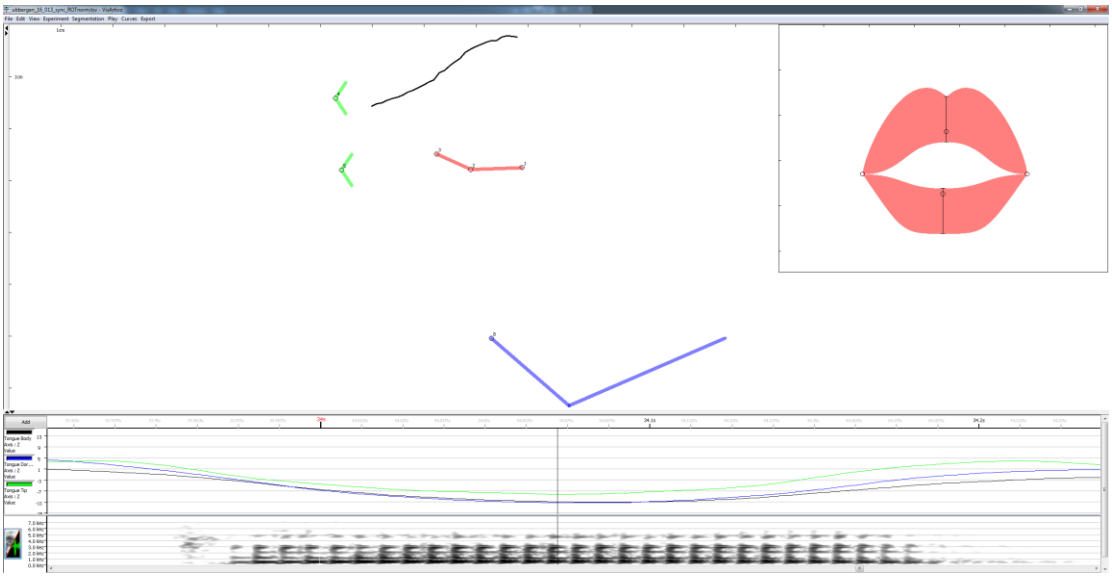

Figure 2. Visualization using VisArtico (Ouni, 2012) of the type of data collected. The top-right inset shows a frontal view of the mouth on the basis of two lip sensors. The top part shows a schematic representation of 2 lip sensors, 3 tongue sensors and 1 jaw sensor. An approximation of the palate of the speaker is also shown. The bottom part shows the trajectories in the inferior-superior dimension for the three tongue sensors. Below those trajectories, the spectral plot is shown for the pronunciation of the standard Dutch CVC sequence [tat].

[2]Besides the three tongue sensors, we also glued three sensors to the lips and attached two sensors to the jaw. For the purpose of this study, however, we only focus on data from the three tongue sensors.

The experiment was divided into two parts. In the first part, participants had to name 70 images (e.g., the image of a ball) in their own dialect (repeated twice, in random order), presented on a computer screen. To familiarize the participants with the images and to make sure they knew what each image depicted, they were asked to name each image in their local dialect once before the sensors were attached. In case the participant failed to use the correct word, he or she was corrected by the experimenter. In the second part, participants had to read 27 CVC sequences out loud (C: /t,k,p/, V: /a,i,o/, e.g., [tap]) in *standard Dutch* (this was emphasized during the explanation of this part). Again, each item was pronounced twice and in randomized order. By including both standard Dutch pronunciations and dialect words, we are able to evaluate if common tongue movement trajectories can be observed in both types of speech. A visual impression of the obtained data can be seen in Figure 2.

Articulatory data preprocessing

After collecting all articulatory data, the data for each speaker were manually segmented (acoustically) at the phone level. Tongue movement data which were not associated with a pronunciation of one of the words included in our study were discarded. The duration of each word's pronunciation was time-normalized between 0 (start of the word) and 1 (end of the word) for each speaker. As the tongue sensors were attached to the midline of the tongue, we only included the position in the inferior-superior direction (i.e. tongue height) and the anterior-posterior direction (i.e. posterior position of the tongue) in our analysis. To enable an appropriate comparison between speakers, the positional information was normalized in such a way that 0 in the inferior-superior direction indicated the lowest (i.e. most inferior) point of the three tongue sensors and 1 the highest point (i.e. most superior). Similarly, 0 in the anterior-posterior direction indicated the most anterior position of the three tongue sensors, while 1 in this direction indicated the most posterior position.

Formant extraction

We automatically extracted the first (F1) and second formant (F2) frequencies of the acoustic recording of the vowels in our dataset using the *findformants* function of the *phonTools* R package (Barreda, 2015). This function extracts formants on the basis of the formulas provided in Snell (1993). We extracted the formants for time slices of 10 milliseconds, centered at the time points for which we had articulatory data. As a rough correction of the automatically extracted formants frequencies, we discarded F1 measurements outside of the range 200 – 1000 Hz, and did the same for F2 measurements outside of the range 500 – 3000 Hz. After this step, we normalized the formant frequencies using Lobanov's (1971) z-transformation, as this normalization method was reported by Adank et al. (2004) to be an adequate normalization procedure retaining sociolinguistic variation. Note that no manual verification of the extracted formants was conducted, so the extracted formants will contain some noise.

Data analysis: generalized additive modeling

Since the articulatory trajectories of the individual tongue sensors are clearly nonlinear, we use generalized additive modeling to analyze the data (Hastie & Tibshirani, 1990; Wood, 2006; see Baayen, 2013 for a non-technical introduction). Generalized additive modeling is a flexible regression approach which not only supports linear relationships between the dependent variable and the independent variables, but also nonlinear dependencies and interactions. In this case our dependent variable is the normalized position of the sensor, which we model as a smooth (i.e. nonlinear) function (SF) over normalized time. The smooth function is represented using a thin plate regression spline (Wood, 2003) which models the nonlinearity as a combination of several low level functions (such as a logarithmic function, a linear function, a quadratic function, etc.). There are other types of splines possible, such as cubic regression splines (consisting of a series of third degree polynomials), but thin plate regression splines have better performance and are computationally efficient (Wood, 2003). To prevent overfitting of the data by the SF, generalized cross-validation is used to determine appropriate parameters of the thin plate regression spline during the model-fitting process (Wood, 2006).

As there is clearly much variation in tongue movement associated with speakers and words, any adequate analysis will need to take this into account. Fortunately, the generalized additive modeling procedure implemented in the R package *mgcv* (version 1.8.7) allows for the inclusion of

factor smooths to represent full random effects. These factor smooths (for an example, see Figure 3) are a nonlinear alternative to random intercepts and random slopes in a mixed-effects regression model. Just as random intercepts and slopes (which are required in a model where multiple observations are present per speaker and/or words; Baayen et al., 2008), factor smooths are essential for taking the structural variability associated with individual speakers and words into account and thereby prevent overconfident (i.e. too low) *p*-values.

As in a common (Gaussian) regression model, the residuals (i.e. the difference between the observed and the estimated values) of a generalized additive model (GAM) have to be independent and normally distributed. However, when analyzing time series which are relatively smooth and slow moving (such as the movement of the tongue over time), the residuals will generally be autocorrelated. This means that the residuals at time *t* will be correlated with the residuals at time *t* + 1 (see Figure 4, left). In our case, the autocorrelation present in the residuals is very high at about 0.97 at lag 1. If this autocorrelation is not brought into the model, the *p*-values of the model will be too small. Fortunately, the function *bam* of the *mgcv* package we use to create the GAMs is able to take into account the autocorrelation of the residuals (see Figure 4, right, where the autocorrelation at lag 1 has been reduced to below 0.1), thereby enabling a more reliable assessment of the model fit and the associated *p*-values. Another important benefit of the *bam* function is that it is able to work with large datasets (Wood et al., 2014), such as the data included in this study (about 1.5 million positions for the dialect data: 40 speakers, 3 sensors, 2 axes, 70 words repeated twice, and a duration of about 0.45 seconds, 45 measurement points, per word; the CVC dataset with 27 words contains about 0.5 million positions).

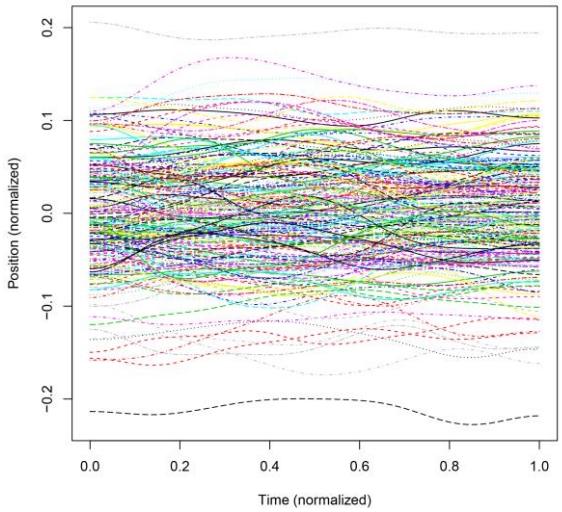

Figure 3. Individual adjustments to the general tongue movement trajectories. As the average of these adjustments is approximately 0 (i.e. centered), both positive and negative adjustments are possible.

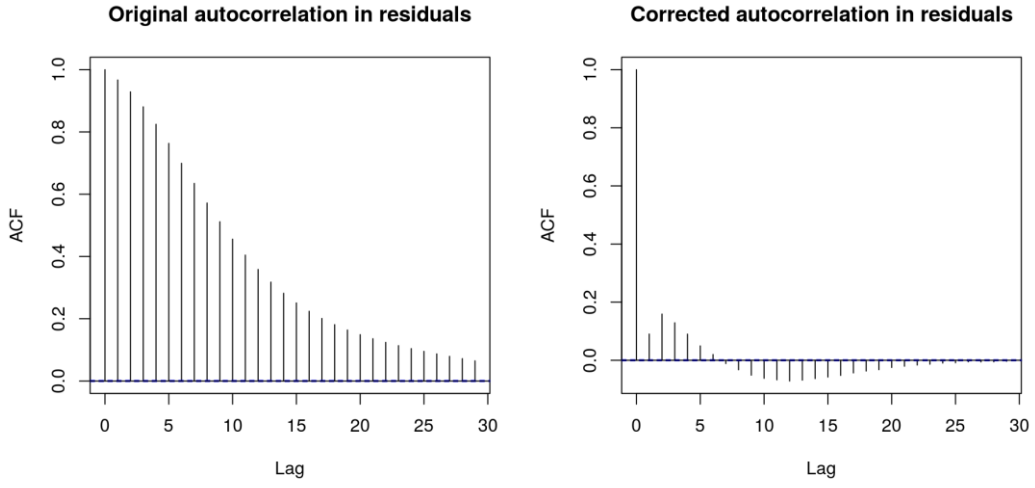

Figure 4. Autocorrelation in the residuals. Left: without correction, right: after correction.

Generalized additive modeling has been used in articulography before (Tomaschek et al., 2013, 2014; Wieling et al., 2015). Furthermore, the method has been applied to language variation research (Wieling et al., 2011 and Wieling et al., 2014), and to model nonlinear patterns across time of brain signals (e.g., Tremblay & Baayen, 2010; Meulman et al., submitted) or gaze data (Van Rij et al., forthcoming).

Reproducibility

To facilitate reproducibility and the use of the methods illustrated in this study, the data, methods and results are available as a paper package stored at the Mind Research Repository (http://openscience.uni-leipzig.de) and the first author's website.

Results

As an illustration of the generalized additive modeling approach, Figure 5 shows the tongue movement trajectories in the oral cavity as measured by the three tongue sensors during the pronunciation of two dialect words: *taarten*, 'cakes' (generally pronounced [tɔːtn] in Ter Apel and [tœʁtə] in Ubbergen), and *boor*, 'drill' (generally pronounced [bʊːr] in Ter Apel and [bʊːʁ] in Ubbergen), as well as two CVC sequences in standard Dutch, *taat*, [tat] and *poop*, [pop]. The red and blue dots in the graph indicate the measured tongue positions of both groups. The red (dark) curves indicate the fitted tongue trajectories of the speakers in Ubbergen for word-specific models, whereas the (lighter) blue curves are linked to the speakers in Ter Apel. The relative lightness of each curve visualizes the time course from the beginning of the word (darkest) to the end of the word (lightest). Clearly the articulations for *taarten* are markedly different for the two groups, whereas the articulations for *boor* are much more similar. In addition, the pronunciations for *taat* show a greater distinction between the two speaker groups than the pronunciations for *poop*. A general pattern across all four graphs, however, is that the speakers from Ubbergen appear to have more anterior tongue positions than those from Ter Apel.

The fitted trajectories were obtained by creating a single GAM for each of the four words for each of the three sensors. In the GAM specification, a different SF was fitted for each group. The command to fit such a model for a single word (simplified: only for a single sensor in a single dimension) using the function `bam` of the `mgcv` package is:

```
model = bam(Position ~ s(Time,by=Group) + Group +
                       s(Time,Speaker,bs='fs',m=1), rho=0.97)
```

The interpretation of this GAM specification is that the sensor position is predicted on the basis of a nonlinear pattern across (normalized) time per group (Ter Apel vs. Ubbergen: `s(Time,by=Group)`), while simultaneously taking into account the speaker-related variation via a factor smooth (the `bs='fs'` block; `m=1` limits the wigglyness of the curve per speaker, which is suitable for these nonlinear random effects). The rho value (in these example specifications fixed at 0.97) indicates the amount of autocorrelation in the residuals which needs to be taken into account (see explanation, above). The linear contrast between the two groups (`Group`) is added to the model as the smooth functions are centered and thus unable to model a constant difference between the two groups.

To see at which points the trajectories differ significantly from each other, confidence intervals are needed. These can readily be extracted from the fitted GAM. Figure 6, visualizing the resulting trajectories and differences for the word *taarten*, shows that the differences in both dimensions are significant across a large part of the time course. While this visualization suggests that the distinction between the two groups is necessary, this should be assessed more formally. There are two approaches for this. The first is fitting a simpler model without the group distinction, and comparing this simpler model to the more complex model having the group distinction to see if the additional complexity is warranted (e.g., by comparing the difference in maximum likelihood scores while taking into account the difference in model complexity). For this we can use the function *compareML* of the R pacakage *itsadug* (version 1.0.1; van Rij et al., 2015). The second approach is to respecify the model in such a way that it does not fit the SFs for the two groups separately, but rather for a single group (i.e. the reference level) and a second smooth function representing the difference between the two groups (i.e. the SF which needs to be added to the SF of the first group to yield the SF

of the second group). The associated *p*-value obtained from the model summary will then indicate if the difference SF is necessary or not. The command to fit this type of model (for a single word) is:

```
diff.model = bam(Position ~ s(Time) + s(Time,by=IsTerApel) +
                            s(Time,Speaker,bs='fs',m=1), rho=0.97)
```

In this case `IsTerApel` is a binary predictor variable equal to 1 for the speakers from Ter Apel and 0 for those from Ubbergen. The SF containing this predictor, `s(Time,by=IsTerApel)`, will be equal to 0 when the binary variable equals 0. This implies that the first smoothing function, `s(Time)`, will be the articulatory trajectory for the Ubbergen group. As the first SF, `s(Time)`, is never equal to 0, this also implies that the SF, `s(Time,by=IsTerApel)`, must be equal to the difference between the Ter Apel and Ubbergen speakers. Since this type of difference SF is not centered as the normal SFs are, no additional contrast between the two groups is necessary. For the visualization in Figure 6, both difference SFs were significant ($p < 0.001$).

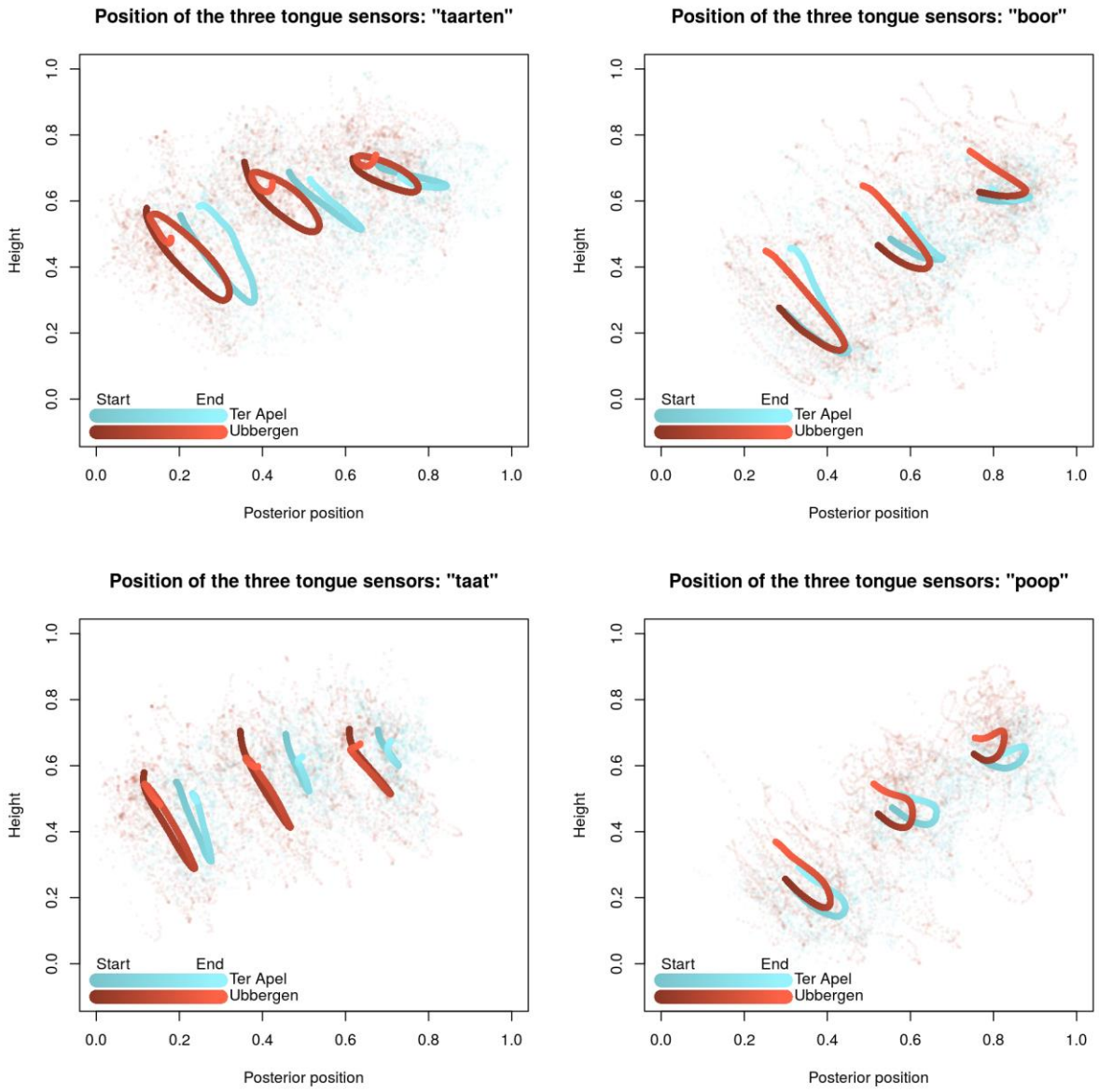

Figure 5. Fitted tongue trajectories (including individual points) of three tongue sensor for the two groups of speakers in two dimensions (posterior position on the *x*-axis, height on the *y*-axis) for two dialect words (up) and two CVC sequences pronounced in Dutch (down). The darkness of the line indicates the time course of the trajectories (dark: start of the pronunciation, light: end of the pronunciation).

While it is certainly insightful to focus on the differences in the pronunciation of individual words, an aggregate analysis is able to provide a more objective view of tongue trajectory differences. In our aggregate models, we simultaneously analyzed the three tongue sensors and two axes for a large set of words. Rather than using a single `s(Time)` for the reference level (Ubbergen) in the simple example above, we now need separate patterns over time for each tongue sensor and axis: i.e. sensor T1 for the inferior-superior axis (i.e. height), sensor T1 for the anterior-posterior axis, sensor T2 for both axes, and sensor T3 for both axes. This can be accomplished by adding a by-parameter distinguishing these six levels (i.e. the interaction between sensor and axis, stored in the variable `SensorAxis`). Similarly, rather than a single SF representing the difference between Ter Apel and Ubbergen (via the use of a binary by-variable), six difference SFs are needed, one for each combination of sensor and axis. Consequently, six binary predictors are created which are equal to 1 for the group of Ter Apel for a specific sensor and axis. For example, the predictor `IsTA.T1.H` equals 1 for the T1 sensor for the inferior-superior axis of the Ter Apel group, while `IsTA.T3.P` is equal to 1 for the T3 sensor for the anterior-posterior axis of the Ter Apel group. Similarly as for the other predictors, the speaker-related variability must also be allowed to vary for each of the six combinations of sensors and axes. This can be achieved by creating a new predictor `SpeakerSensorAxis` representing the interaction between the three predictors `Speaker`, `Sensor`, `Axis` and using this predictor in the factor smooth. Given that we are now aggregating over a large set of words, we also need to take into account the variability per word (per sensor and axis separately) via a factor smooth. The specification of this model is as follows:

```
model = bam(Position ~ s(Time,by=SensorAxis) + SensorAxis +
                       s(Time,by=IsTA.T1.H) + s(Time,by=IsTA.T1.P) +
                       s(Time,by=IsTA.T2.H) + s(Time,by=IsTA.T2.P) +
                       s(Time,by=IsTA.T3.H) + s(Time,by=IsTA.T3.P) +
                       s(Time,SpeakerSensorAxis,bs='fs',m=1) +
                       s(Time,WordSensorAxis,bs='fs',m=1), rho=0.97)
```

Following this model specification, we created two different large-scale GAMs. The first GAM assessed the tongue trajectories for the two groups of speakers for the 70 dialect words (about 1.5 million positions, taking about 4 hours on a 36-core Intel Xeon E5-2699 v3), while the second GAM focused on the 27 CVC sequences pronounced in standard Dutch (about 500,000 positions, taking about 1 hour on the same server). We did not include the data of both the standard language and the dialects in a single GAM, as the items in both sets are not adequately comparable. The standard Dutch pronunciations always consist of a CVC sequence, whereas this is not the case for the dialect words.

The results of the model for the dialect words are shown in Tables 1 and 2. The explained variance of the model is equal to 86.1%, due mainly to the inclusion of the factor smooths per speaker and word. The parametric part of the model shown in Table 1 simply compares the posterior position of the T3 sensor to the height of the T3 sensor and the height and posterior position of the other sensors. While the comparison between height and posterior position is not informative as such, these comparisons are required as the model includes both dimensions simultaneously. As expected, Table 1 shows that the position of the T2 (middle) sensor is more anterior (i.e. negative) than the T3 sensor, while the T1 sensor is more anterior than the T3 and T2 sensor. Table 2 gives some information about the SFs used and shows (in lines 7 to 12) that the difference SFs between the two groups are significant with the exception of the posterior position difference for the T3 sensor ($p = 0.219$), and the posterior position difference for the T1 sensor ($p = 0.051$). However, to be able to interpret these SFs, visualization is essential.

Figures 7 and 8 visualize the results with respect to the posterior position and height and clearly show that during the pronunciation of the dialect words the tongue of the speakers in Ter Apel is generally positioned more posterior than the tongue of the speakers in Ubbergen (non-significantly so for the T1 and T3 sensor).

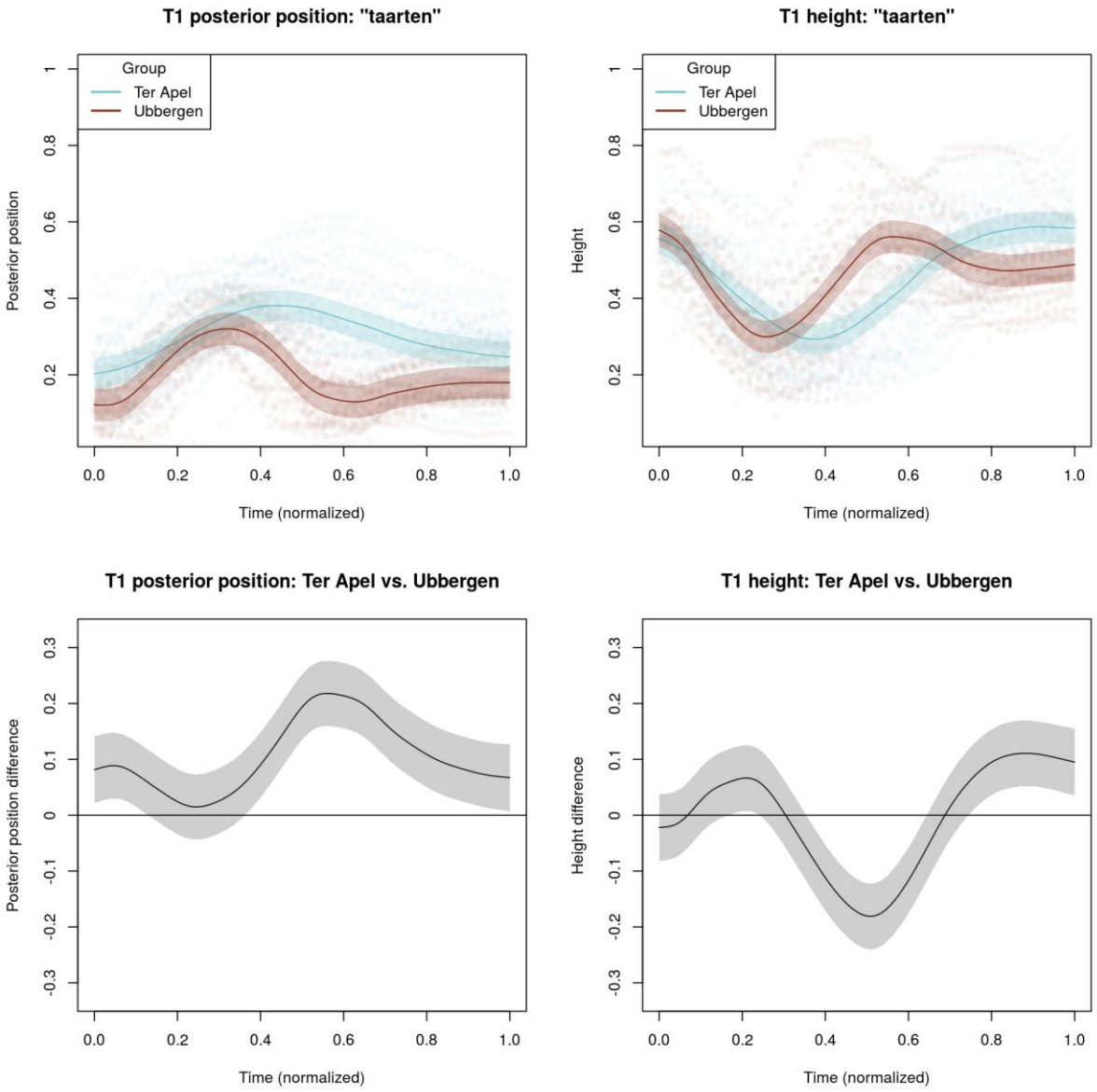

Figure 6. T1sensor and sensor difference trajectories for the word *taarten* in the anterior-posterior dimension (left) and the height dimension (right) for both groups. The upper graphs show the trajectories per group including 95% confidence bands together with the individual points. The lower graphs show the difference between the two groups including confidence bands extracted from the fitted GAM (which took the individual variation and autocorrelation in the residuals into account).

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (T3 posterior position)	0.696	0.014	48.4	< 0.001
T2 posterior position vs. T3 posterior position	-0.241	0.020	-11.8	< 0.001
T1 posterior position vs. T3 posterior position	-0.470	0.020	-23.0	< 0.001
T3 height vs. T3 posterior position	0.003	0.020	0.1	0.900
T2 height vs. T3 posterior position	-0.121	0.020	-5.9	< 0.001
T1 height vs. T3 posterior position	-0.278	0.021	-13.5	< 0.001

Table 1. Parametric coefficients of the model on the basis of all dialect words, all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height).

Smooth Functions (SFs)	edf	*F*-value	*p*-value
s(Time) : T3 posterior position	4.1	1.7	0.130
s(Time) : T2 posterior position	7.9	4.3	< 0.001
s(Time) : T1 posterior position	10.9	6.4	< 0.001
s(Time) : T3 height	8.9	7.7	< 0.001
s(Time) : T2 height	11.5	14.1	< 0.001
s(Time) : T1 height	16.2	14.5	< 0.001
s(Time) : T1 posterior position difference SF	5.7	2.0	0.051
s(Time) : T1 height difference SF	17.1	9.9	< 0.001
s(Time) : T2 posterior position difference SF	4.1	2.9	0.013
s(Time) : T2 height difference SF	13.8	3.5	< 0.001
s(Time) : T3 posterior position difference SF	2.0	1.5	0.219
s(Time) : T3 height difference SF	10.1	2.5	0.003
s(Time, Speaker) [factor smooth]	1949.5	39.4	< 0.001
s(Time, Word) [factor smooth]	3722.4	121.9	< 0.001

Table 2. SF terms of the model on the basis of all dialect words, all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height). The first 6 lines show the SFs for the reference level (Ubbergen), whereas lines 7 to 12 represent difference SFs (comparing Ter Apel to Ubbergen). The edf column indicates the estimated degrees of freedom, which is a measure to reflect SF complexity. The maximum allowed SF complexity was 19 edf (enforced by setting the k-parameter of each SF to 20), and this seems to be sufficiently high as none of the SFs have an edf close to 19. The *p*-value assesses if the SF is significantly different from 0. The final two lines show the factor smooths per speaker and word.

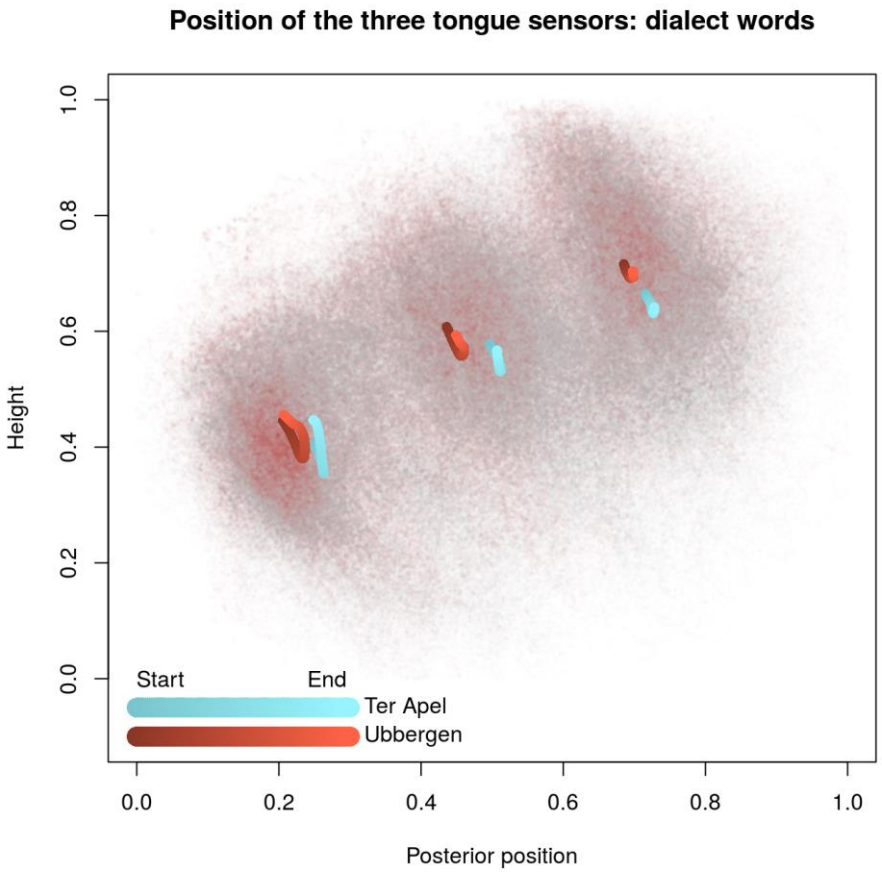

Figure 7. Aggregate fitted tongue trajectories of three tongue sensors (T1: front, T2: middle, T3: back) for the two groups of speakers in two dimensions (posterior position on the *x*-axis, height on the *y*-axis) for all 70 dialect words. The darkness of the lines indicates the time course of the trajectories (dark: start of the pronunciation, light: end of the pronunciation). The difference in anterior-posterior position is significant for T2 , while the height differences are significant for all sensors (see Table 2).

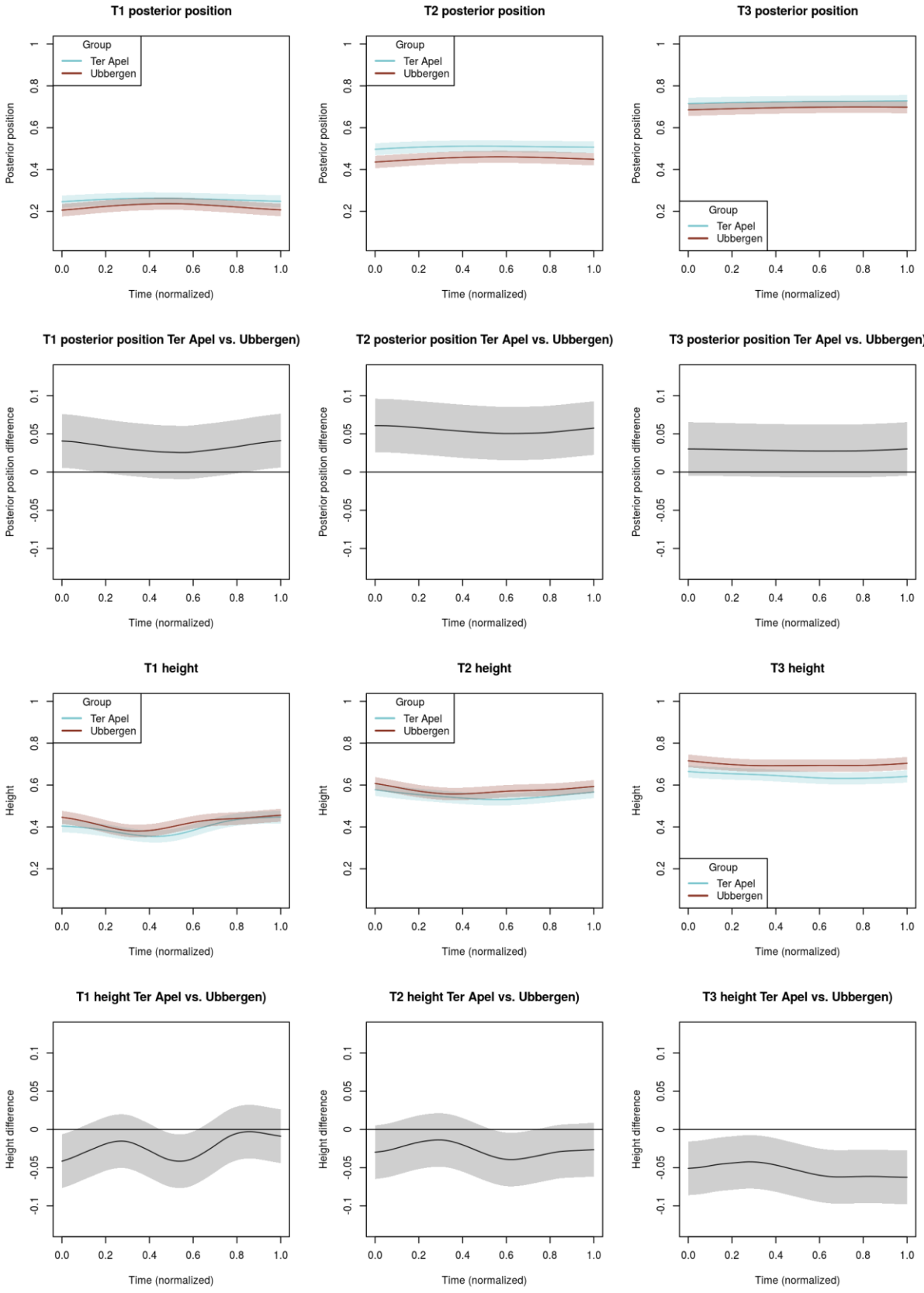

Figure 8. Graphs in row 1: tongue sensor trajectories aggregated over all 70 dialect words in the anterior-posterior dimension for both groups. Graphs in row 2: differences between tongue sensor trajectories in the anterior-posterior dimension. The difference is significant ($p < 0.05$) for the middle tongue sensor (T2), but not for the frontal tongue sensor (T1; $p = 0.051$), or the back tongue sensor (T3; $p = 0.219$, see Table 2). Graphs in row 3 and 4 show the corresponding results for height. All of the height differences are significant ($p < 0.01$).

Similar to the dialect words, Table 3 and 4 show the results for the model on the basis of the standard Dutch CVC sequences (explaining 92.2% of the variance). Interestingly, Figures 9 and 10 show a similar pattern with respect to the posterior position compared to Figures 7 and 8, with speakers from Ter Apel having their tongue more posterior than the speakers from Ubbergen. Even though the target pronunciation (standard Dutch) is the same in this case, the tongue differences in this dimension are comparable to those observed on the basis of dialectal speech.

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (T3 posterior position)	0.666	0.018	37.3	< 0.001
T2 posterior position vs. T3 posterior position	-0.237	0.025	-9.4	< 0.001
T1 posterior position vs. T3 posterior position	-0.452	0.025	-17.9	< 0.001
T3 height vs. T3 posterior position	0.079	0.025	3.2	0.002
T2 height vs. T3 posterior position	-0.088	0.025	-3.5	< 0.001
T1 height vs. T3 posterior position	-0.283	0.025	-11.2	< 0.001

Table 3. Parametric coefficients of the model on the basis of all Dutch CVC sequences, all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height).

Smooth Functions (SFs)	edf	*F*-value	*p*-value
s(Time) : T3 posterior position	13.9	4.0	< 0.001
s(Time) : T2 posterior position	14.1	4.7	< 0.001
s(Time) : T1 posterior position	15.1	6.0	< 0.001
s(Time) : T3 height	15.3	12.0	< 0.001
s(Time) : T2 height	16.5	24.1	< 0.001
s(Time) : T1 height	17.2	22.3	< 0.001
s(Time) : T1 posterior position difference SF	13.4	3.8	< 0.001
s(Time) : T1 height difference SF	13.8	2.8	< 0.001
s(Time) : T2 posterior position difference SF	11.2	3.5	< 0.001
s(Time) : T2 height difference SF	11.8	3.8	< 0.001
s(Time) : T3 posterior position difference SF	10.0	2.6	0.002
s(Time) : T3 height difference SF	2.0	0.9	0.396
s(Time, Speaker)	1910.9	28.2	< 0.001
s(Time, Word)	1408.8	144.7	< 0.001

Table 4. Smooth terms of the model on the basis of all Dutch CVC sequences, all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height). The first 6 lines show the smooths for the reference level (Ubbergen), whereas lines 7 to 12 represent the difference SFs (comparing Ter Apel to Ubbergen). The edf column indicates the estimated degrees of freedom, which is a measure reflecting the SF complexity. The maximum allowed SF complexity was 19 edf (enforced by setting the k-parameter of each SF to 20), and this seems to be sufficiently high as none of the SFs have an edf close to 19. The *p*-value assesses if the SF is significantly different from 0. The final two lines show the factor smooths per speaker and word.

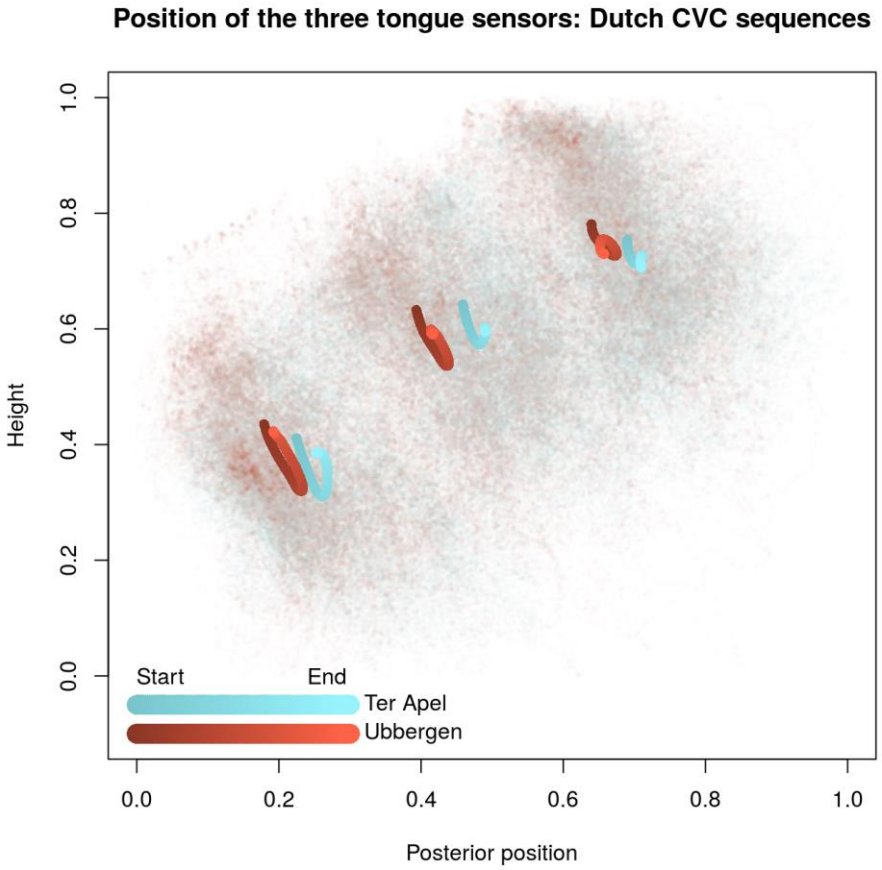

Figure 9. Aggregate fitted tongue trajectories of three tongue sensors (T1: front, T2: middle, T3: back) for the two groups of speakers in two dimensions (posterior position on the *x*-axis, height on the *y*-axis) for all 27 standard Dutch CVC sequences. The darkness of the lines indicates the time course of the trajectories (dark: start of the pronunciation, light: end of the pronunciation). The differences in posterior position and height are all significant ($p < 0.01$), except for the T3 height difference (see Table 4).

Comparison to linear discriminant analysis

Since generalized additive modeling is a relatively new technique, especially when applied to articulatory data (see Tomaschek et al., 2013 and 2014), we have also analyzed the data using another technique, namely linear discriminant analysis (LDA).[3] In LDA an item's class (in our case the group of the speaker) is predicted on the basis of a set of numerical predictors (in our case the normalized height and posterior position for the three tongue sensors). For both the dialect data and the CVC data, we created five different LDAs using segment-specific positions (i.e. for /a/, /i/, /o/, /k/ and /t/). All ten LDAs showed significant group mean differences (all p's < 0.001) generally in line with the global position differences shown in Figures 8 and 10. Thus, for both datasets the sensor positions were more posterior and lower for the speakers from Ter Apel than for the speakers from Ubbergen. The probability of correctly classifying the group of the speaker on the basis of the tongue position (on the basis of the three sensors) at a certain time point ranged between 67% and 83% (see Table 5). In sum, the LDA analysis showed that the tongue position (in terms of the height and posterior position of the three tongue sensors) during the pronunciation of a single segment is useful for predicting from which dialect region a speaker originates. These results are in line with the results on the basis of the generalized additive modeling approach, which also showed clear differences between the groups.

[3] Note that LDA is not entirely appropriate for data with repeated measures (Lix and Sajobi, 2010). In addition, LDA requires observations to be independent, which assumption is violated in this dataset where each individual speaker contributes many tongue positions. Consequently, the LDA approach may be anti-conservative when applied to this dataset. A repeated-measures LDA approach would be more appropriate, but to our knowledge no such procedure is implemented in R.

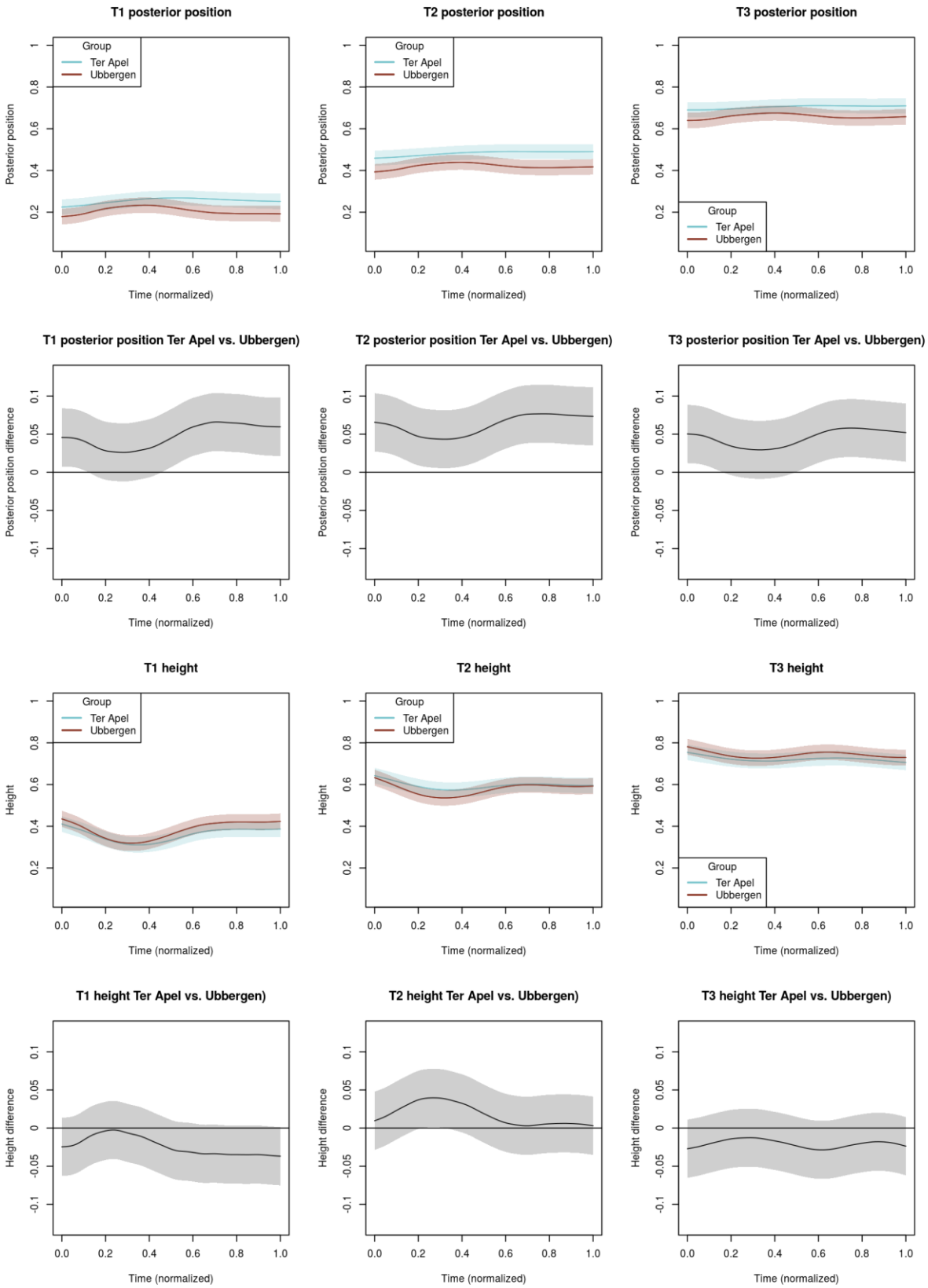

Figure 10. Graphs in row 1: tongue sensor trajectories aggregated over all 27 standard Dutch CVC sequences in the anterior-posterior dimension for both groups. Graphs in row 2: differences between tongue sensor trajectories. All difference SFs are significant ($p < 0.01$). Graphs in row 3 and 4 show the corresponding results for height. All difference SFs are significant ($p < 0.01$), except for the T3 height difference ($p = 0.396$). Significance indicates that the probability of the difference over time being equal to 0 *across the whole time* span is smaller than 0.05. Consequently, the nonlinear difference is necessary to improve the model fit, even though the difference at each single time point may be not significantly different from 0.

Dataset	/a/	/i/	/o/	/k/	/t/
Dialect words	68%	71%	73%	73%	83%
CVC sequences	71%	71%	67%	69%	81%

Table 5. Speaker group classification accuracy on the basis of the height and posterior position of the three tongue sensors.

Comparison with formant-based patterns

It is generally assumed that tongue height correlates with F1, and posterior tongue position with F2 (Stevens, 1998). In line with this, the average correlation between the normalized height of the tongue sensors and the normalized F1 frequency for the dialect words was r = -0.27 (all p's < 0.001), for the CVC sequences (containing only three different vowels) the correlation increased to r = -0.49 (all p's < 0.001). The correlations are negative as a higher F1 is related to a lower tongue position. When looking at the normalized posterior position of the tongue sensors, the average correlation with the normalized F2 frequency was r = -0.51 (all p's < 0.001) for the dialect words and r = -0.67 (all p's < 0.001) for the CVC sequences. Again, the correlations are negative as a higher F2 is related to a less posterior tongue position. Figures 11 and 12 visualize the scatter plots between the formants and the tongue sensor positions for the dialect words and the CVC sequences, respectively.

In the previous section, we observed a more posterior tongue position for both the dialect words as well as the CVC sequences for the speakers from Ter Apel versus those from Ubbergen. Furthermore, for the dialect words, the speakers from Ter Apel had lower tongue height than those from Ubbergen. Consequently, we would expect lower F2 values for the speakers from Ter Apel compared to those from Ubbergen, and higher F1 values, but only for the dialect words. However, this is not what we observe. The normalized F1 values were higher for the speakers from Ter Apel for the CVC words, but lower for the dialect words (all $|t|$'s > 5 in mixed-effects regression models including speaker as a random-effect factor). Similarly, the normalized F2 values were higher for the speakers from Ter Apel for the CVC words, but lower for the dialect words. Thus, only the F2 difference for the dialect words is in line with what we would expect on the basis of the articulatory results.

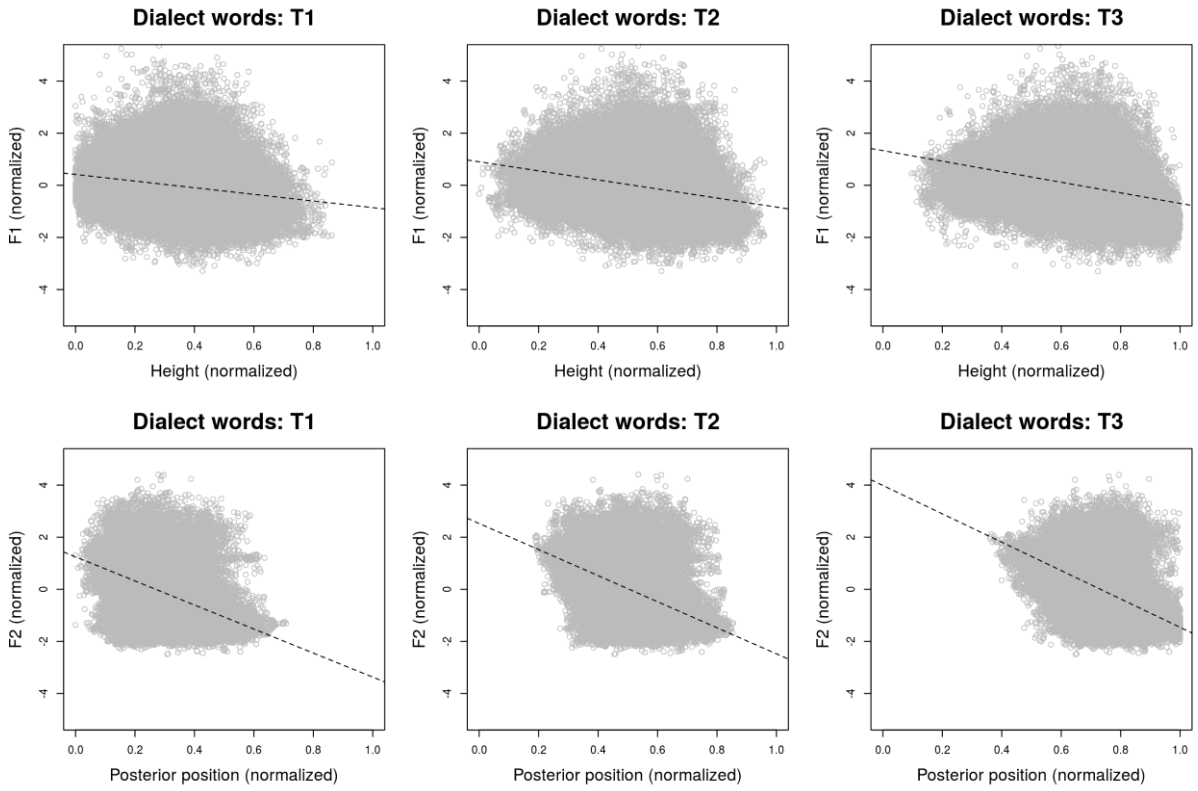

Figure 11. The three graphs in row 1 visualize the relation for the dialect words between normalized height and normalized F1 for the three tongue sensors. Those in row 2 show the same for normalized posterior position and normalized F2.

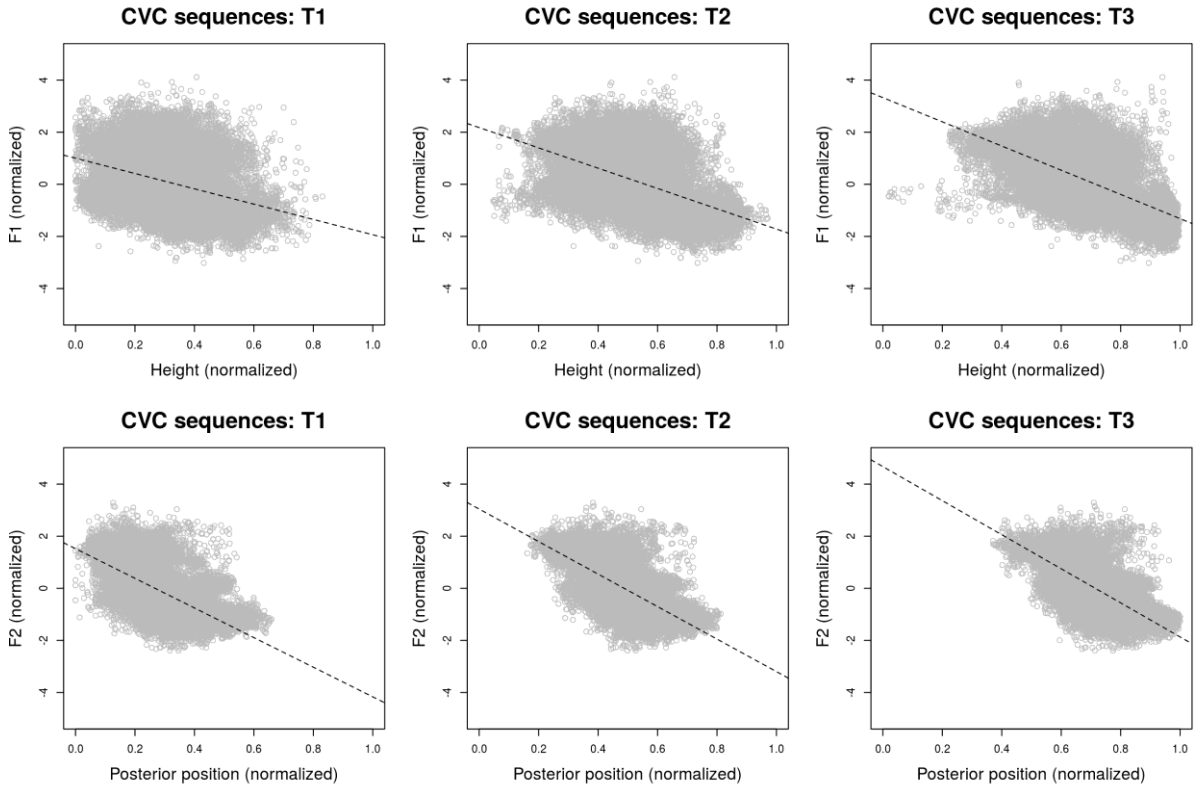

Figure 12. The three graphs in row 1 visualize the relation for the CVC sequences between normalized height and normalized F1 for the three tongue sensors. Those in row 2 show the same for normalized posterior position and normalized F2.

Discussion

In this study we have illustrated the use of articulatory data for the purpose of studying dialect variation. We identified a structural difference in the position of the tongue between the two groups of speakers, with more anterior positions of the tongue for the speakers from Ubbergen in the southern half of the Netherlands compared to the speakers from Ter Apel in the northern half of the Netherlands. This result contrasts with previous findings on Dutch dialects of Adank et al. (2007) who did not find a difference in F2 for two (corresponding) groups using only a single formant measurement for monophthongs. However, Van der Harst et al. (2014) show that a dynamic approach using acoustic vowel information (F1 and F2) measured across multiple time points does help in uncovering regional differences. While we also discovered regional differences on the basis of formants measured at multiple time points, these did not line up with the articulatory results in line with our expectations. While this may have been caused by noise in the automatically obtained formant frequencies or the (unobserved) parasagittal position of the tongue, there is also no one-to-one correspondence between F1/F2 and height/posterior position of the tongue (despite these values being frequently interpreted as such since Bell, 1867). The clear discrepancy between the patterns observed on the basis of articulatory data versus those on the basis of acoustic data, emphasizes the need and use for articulatory data in studies investigating language variation.

Our findings are well interpretable in the context of articulatory settings (Honikman, 1964; Laver, 1978). Honikman (1964, p. 73) defined the articulatory setting as "the overall arrangement and manoeuvring of the speech organs necessary for the facile accomplishment of natural utterance". Given that the speakers from Ter Apel showed a tongue position which was more posterior than the speakers from Ubbergen, both when pronouncing words in their own dialect and even more so in standard Dutch, this suggests that there are distinct articulatory settings for the two dialects, causing distinguishable accents when pronouncing standard Dutch. Whereas distinct articulatory settings have been identified for individual languages, such as English and French (Honikman, 1964; Gick et al., 2004), no articulatory differences have been previously reported at the dialectal level.

The generalized additive modeling approach proposed here results in a model of tongue movement over time, while taking into account individual and word-related variability, as well as autocorrelation in the residuals. While we have not used this here, the generalized additive model may also be used to determine speed and acceleration of the fitted trajectories. The approach complements other approaches used to analyze articulatory data over time, such as dynamic time warping (Sakoe & Chiba, 1978), functional data analysis (e.g., Lucero, Munhall, Gracco & Ramsay, 1997), or cross-recurrence analysis (Lancia, Fuchs and Tiede, 2014). These methods generally separate amplitude variability from phase variability when comparing articulatory trajectories. The method we propose, however, is particularly suitable when articulatory trajectories need to be compared at a higher level of aggregation for a large number of speakers. Furthermore, our approach is able to take into account individual variation, and correct for autocorrelation in the residuals.

We have shown that the results using generalized additive modeling are in line with those using linear discriminant analysis. However, there are clear benefits of generalized additive modeling over linear discriminant analysis. First, it is not necessary to create separate analyses for each individual segment, and second, the GAM analysis takes into account individual variation by treating speaker as a random-effect factor.

Whereas the two dialects studied here show clear pronunciation differences which can usefully be studied from an acoustic or transcription-based dialectometric perspective, the aggregate articulatory perspective put forward in this study revealed interesting results, which were markedly different when taking an acoustic perspective. This further shows that articulatory data is not only an essential component in an integrated account of socially-stratified variation (Lawson et al., 2011), but also for regionally-stratified variation.

Acknowledgements

This work is part of the research program *Investigating language variation physically*, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO) via a Rubicon grant awarded to Martijn Wieling. Furthermore, this work has benefitted from funding of the Alexander von Humboldt Professorship awarded to R. Harald Baayen. We thank Dankmar Enke, Matthias Villing, Lea Hofmaier, and Amber Nota for help in segmenting the acoustic data.

References

Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099-3107.

Adank, P., van Hout, R., & Van de Velde, H. (2007). An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *The Journal of the Acoustical Society of America*, *121*(2), 1130-1141.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Baayen, R. H. (2013). Multivariate Statistics. In Podesva, R. and D. Sharma, D. (eds.), *Research Methods in Linguistics*, pp. 337-372. Cambridge: Cambridge University Press.

Barreda, S. (2015). phonTools: Functions for phonetics in R. R package version 0.2-2.1.

Bell, A. M. (1867). *Visible Speech: The Science of Universal Alphabetics Or, Self-interpreting Physiological Letters, for the Writing of All Languages in One Alphabet*. Simpkin, Marshall & Company.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*(1), 111-140.

Clopper, C. G., & Paolillo, J. C. (2006). North American English vowels: A factor-analytic perspective. *Literary and linguistic computing*, *21*(4), 445-462.

Corneau, C. (2000). An EPG study of palatalization in French: Cross-dialect and inter-subject variation. *Language Variation and Change*, *12*(1), 25-49.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variancea). *The Journal of the Acoustical Society of America*, *120*(1), 407-415.

Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, *54*(1), 1-21.
Gick, B., Wilson, I., Koch, K., & Cook, C. (2005). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, *61*(4), 220-233.
Goeman, A. (1999). *T-deletie in Nederlandse dialecten. Kwantitatieve analyse van structurele, ruimtelijke en temporele variatie*. Holland Academic Graphics.
Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. CRC Press.
Heeringa, W. (2004). Measuring Dialect Pronunciation Differences using Levenshtein Distance. PhD thesis, University of Groningen.
Honikman, B. (1964). Articulatory Settings. In D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, & J.L.M. Trim (Eds.), *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961*, pp. 73-84. London: Longmans, Green & Co. Ltd.
Hoole, P., & Zierdt, A. (2010). Five-dimensional articulography. *Speech motor control: New developments in basic and applied research*, 331-349.
Hoole, P., & Nguyen, N. (1999). Electromagnetic articulography. *Coarticulation–Theory, Data and Techniques, Cambridge Studies in Speech Science and Communication*, 260-269.
Kerswill, P., & Wright, S. (1990). The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change*, *2*(03), 255-275.
Koos, B., Horn, H., Schaupp, E., Axmann, D., & Berneburg, M. (2013). Lip and tongue movements during phonetic sequences: analysis and definition of normal values. *The European Journal of Orthodontics*, *35*(1), 51-58.
Labov, William. (1980). The social origins of sound change. In Labov, William (ed.), *Locating language in time and space*. New York: Academic Press.
Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Vol. 1). US Regional Survey.
Lancia, L., Fuchs, S., & Tiede, M. (2014). Application of Concepts From Cross-Recurrence Analysis in Speech Production: An Overview and Comparison With Other Nonlinear Methods. *Journal of Speech, Language, and Hearing Research*, 1-16.
Laver, J. (1978). The concept of articulatory settings: an historical survey.*Historiographia Linguistica*, *5*(1-2), 1-14.
Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic/r/in Scottish English. *Journal of Sociolinguistics*, *15*(2), 256-268.
Leinonen, T. N. (2010). *An acoustic analysis of vowel pronunciation in Swedish dialects*. PhD thesis, Rijksuniversiteit Groningen.
Lix, L. M., & Sajobi, T. T. (2010). Discriminant analysis for repeated measures data: a review. *Frontiers in psychology*, *1*.
Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*, 606–608.
Lucero, J. C., Munhall, K. G., Gracco, V. L., & Ramsay, J. O. (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research*, *40*(5), 1111-1117.
Meulman, N., Wieling, M., Sprenger, S.A., Stowe, L.A., & Schmid, M.S. (submitted). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them.
Nerbonne, J., Heeringa, W., Van den Hout, E., Van de Kooi, P., Otten, S., & Van de Vis, W. (1996). Phonetic distance between Dutch dialects. In: Durieux, G., Daelemans, W., & Gillis, S. (eds.), CLIN VI: Proceedings of the Sixth CLIN Meeting, Antwerp, pp. 185–202.
Ouni, S., Mangeonjean, L., & Steiner, I. (2012), VisArtico: a visualization tool for articulatory data, Proceedings of Interspeech 2012, September 9-13, 2012, Portland, OR, USA.
Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, *92*(6), 3078-3096.
Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175-184.

Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear/l/and dark/l: evidence from two Catalan dialects. *Journal of the International Phonetic Association*, *35*(01), 1-25.

Recasens, D., & Espinosa, A. (2007). An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association*, *37*(02), 143-172.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America*, *125*(4), 2288-2298.

Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production*. Oxford University Press.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43-49.

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, *31*(1), 26-35.

Scobbie, J.M. & K. Sebregts (2010). Acoustic, articulatory and phonological perspectives on allophonic variation of /r/ in Dutch. In: Folli, R. & C. Ulbrich (eds.), Interfaces in Linguistics: New Research Perspectives. Oxford: Oxford University Press.

Stevens, K. N. 1998. Acoustic phonetics. Cambridge, MA: MIT Press.

Sweet, H. (1888). *A history of English sounds from the earliest period: with full word-lists*. Clarendon Press.

Tabain, M. (2013). Research methods in speech production. In: Jones, M. & Knight, R.-A. (eds) *Bloomsbury Companion to Phonetics*, London: Bloomsbury, pp. 39-56.

Tomaschek, F., Tucker, B. V., Wieling, M., & Baayen, R. H. (2014). Vowel articulation affected by word frequency. Proceedings of the 10th ISSP, Cologne, pp. 429-432.

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. Proceedings of the 14th Interspeech, Lyon, pp. 1302-1306.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, 151-173.

Van Rij, J., Wieling, M., Baayen, R.H., van Rijn, H. (2015). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package, version 1.0.1.

Van Rij, J., Hollebrandse, B., & Hendriks, P. (forthcoming). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In: Holler, A., Goeb, C., & Suckow, K. (eds.) *Experimental Perspectives on Anaphora Resolution. Information Structural Evidence in the Race for Salience*.

Van der Harst, S., Van de Velde, H., & Van Hout, R. (2014). Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language Variation and Change*, *26*(2), 247-272.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-Lehouillier, H., Vatikiotis-Bateson. E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech Language and Hearing Research*, *48*, 543-553.

Wieling, M., Heeringa, W., & Nerbonne, J. (2007). An aggregate analysis of pronunciation in the Goeman-Taeldeman-Van Reenen-Project data. *Taal en Tongval*, *59*, 84–116.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R.H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669-692.

Wieling, M., & Nerbonne, J. (2011). Bipartite spectral graph partitioning for clustering dialect varieties and detecting their linguistic features. *Computer Speech and Language*, *25*(3), 700-715.

Wieling, M., & Nerbonne, J. (2015). Advances in Dialectometry. *Annual Review of Linguistics*, *1*(1).

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLOS ONE*, *6*(9), e23613.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., & Baayen, R. H.. Investigating dialectal differences using articulography. Proceedings of ICPhS 2015, Glasgow.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95-114.
Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.
Wood, S. N., Goude, Y., & Shaw, S. (2014). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.
Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica*, *64*(2), 94-102.
Yunusova, Y., Green, J. R., & Mefferd, A. (2009). Accuracy assessment for AG500, Electromagnetic Articulograph. *Journal of Speech Language and Hearing Research, 52*, 547-555.